Car Sales Analysis
Group 05 Project

6th December 2024

Submitted To

Professor Thiru Pandian

BUAN 6320.007 – Database Foundations for Business Analytics

Prepared By :

Vedansh Avlani   **- VHA240000**
Natesh Subramanian   **- NXS240009**
Rahul Tamanam   **- RXT240022**
Anushka Gupta   **- AXG240051**
Pinank Nagda   **- PXN141030**
Chanakya Sairam Alluri   **- CXA230019**

Table of Contents

Introduction

I.   Project Description

This project aims to analyze a dataset of car sales using SQL and data visualization tools to generate actionable insights. The dataset includes attributes related to vehicles, their sales performance, and regional market dynamics.

Objectives:

1. Understand trends in car sales, including price appreciation and depreciation.

2. Analyze the impact of factors such as condition, transmission type, and odometer readings on selling prices.

3. Explore regional differences in market performance and price distribution.

4. Identify high-performing vehicle brands and models.

5. Provide actionable insights for both buyers and sellers in the used car market.

II.   About the Dataset

Source:
https://www.kaggle.com/datasets/syedanwarafridi/vehicle-sales-data

- Key Features:
  - Vehicle Details:
    - Make
    - Model
    - Year of Manufacture
    - Body Type
    - Condition
    - Transmission Type
    - Odometer Reading

  - Sales Details:
    - Sale Date
    - Selling Price
    - MMR

  - Regional Information:
    - State
    - Country

- Dataset Size:
  - Number of records: 558837 records
  - Number of columns: 16 columns

III. Data Loading Concept

- •A MySQL database named ap was used to store the dataset.
- •The SQL file included CREATE TABLE commands to define the schema for the data.
- •The SQL file included multiple INSERT INTO commands to load the data into the newly created table.
- •The SQL script was executed using MySQL Workbench or the MySQL Command Line Interface.

```sql
1   CREATE DATABASE  IF NOT EXISTS `ap` /*!40100 DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci */ /*!80016 DEFAULT ENCRYPTION='N' */;
2   USE `ap`;
3
4   /*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
5   /*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
6   /*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
7   /*!50503 SET NAMES utf8 */;
8   /*!40103 SET @OLD_TIME_ZONE=@@TIME_ZONE */;
9   /*!40103 SET TIME_ZONE='+00:00' */;
10  /*!40014 SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0 */;
11  /*!40014 SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0 */;
12  /*!40101 SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='NO_AUTO_VALUE_ON_ZERO' */;
13  /*!40111 SET @OLD_SQL_NOTES=@@SQL_NOTES, SQL_NOTES=0 */;
14
15  DROP TABLE IF EXISTS `vehicle_sales`;
16  /*!40101 SET @saved_cs_client     = @@character_set_client */;
17  /*!50503 SET character_set_client = utf8mb4 */;
18  CREATE TABLE `vehicle_sales` (
19    `year` int DEFAULT NULL,
20    `make` varchar(50) DEFAULT NULL,
21    `model` varchar(50) DEFAULT NULL,
22    `trim` varchar(50) DEFAULT NULL,
23    `body` varchar(50) DEFAULT NULL,
24    `transmission` varchar(20) DEFAULT NULL,
25    `vin` varchar(17) DEFAULT NULL,
26    `state` varchar(2) DEFAULT NULL,
27    `condition` int DEFAULT NULL,
28    `odometer` int DEFAULT NULL,
29    `color` varchar(20) DEFAULT NULL,
30    `interior` varchar(20) DEFAULT NULL,
31    `seller` varchar(100) DEFAULT NULL,
32    `mmr` decimal(10,2) DEFAULT NULL,
33    `sellingprice` decimal(10,2) DEFAULT NULL,
34    `saledate` datetime DEFAULT NULL
35  ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
36  /*!40101 SET character_set_client = @saved_cs_client */;
37
38  LOCK TABLES `vehicle_sales` WRITE;
39  INSERT INTO `vehicle_sales` VALUES (2015,'Kia','Sorento','LX','SUV','automatic','5xyktca69fg566472','ca',5,16639,'white','black','kia motors america  inc',20500.00,21500.00,'2014-12-16 00:00:00'),(2015,'Kia','Sorento','LX',
```

IV. Insights

a) Insight 1

Which car brands have shown the most price appreciation over the past few years?

SQL query:

```
WITH price_trend AS (

    SELECT make, year, AVG(sellingprice) AS avg_price

    FROM vehicle_sales

    GROUP BY make, year

)

SELECT make, (MAX(avg_price) - MIN(avg_price))/MIN(avg_price) AS relative_price_appreciation

FROM price_trend

GROUP BY make

HAVING COUNT(year) > 1

ORDER BY relative_price_appreciation DESC;
```

Visualization of Output:
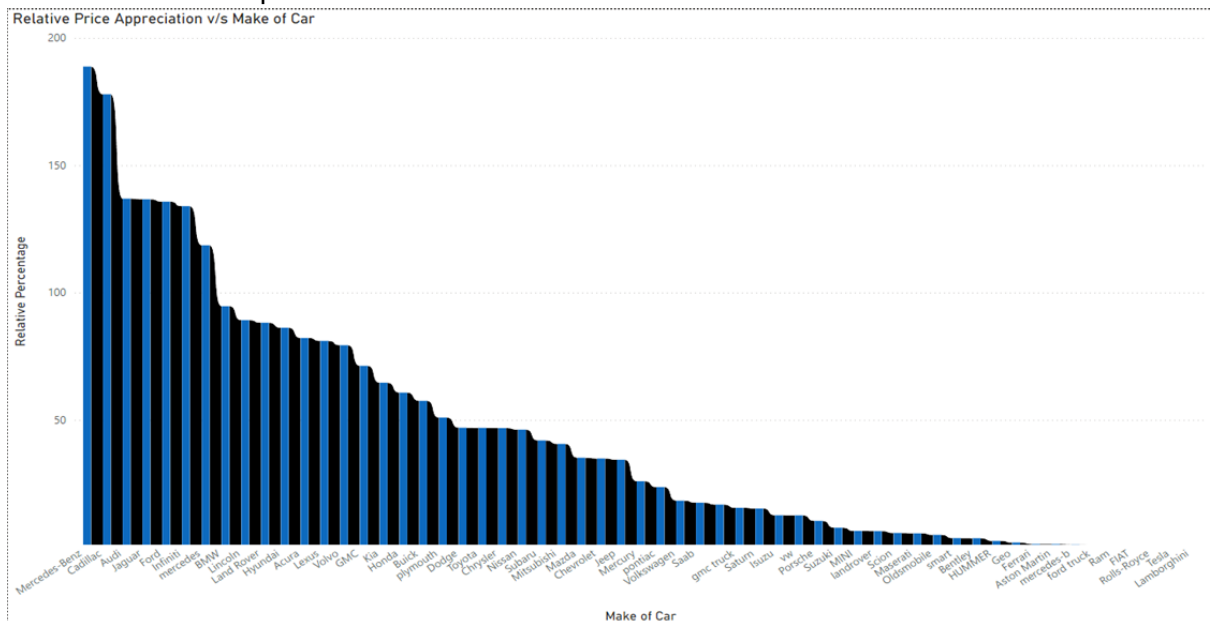
Interpretation of Result:
Mercedes-Benz, Cadillac, and Jaguar have exhibited the highest appreciation in selling prices over time, indicating strong demand and value retention in the market.
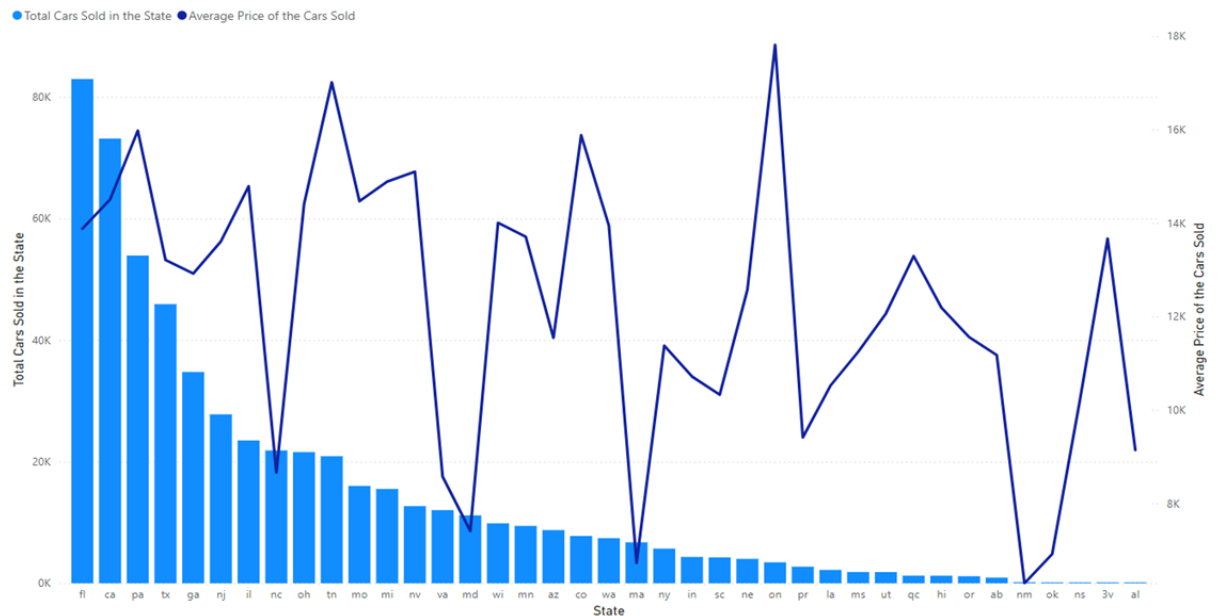
b) Insight 2

What is the overall market performance for used cars in different states?

SQL query:

```
SELECT state, COUNT(*) AS total_sales, AVG(sellingprice) AS average_price
FROM vehicle_sales
GROUP BY state
ORDER BY average_price DESC;
```

Visualization of Output:

Interpretation of Result:

Florida (FL) leads the market in total used car sales, while states like Colorado (CO) and Nebraska (NE) show higher average prices, indicating potential premium markets for used cars.
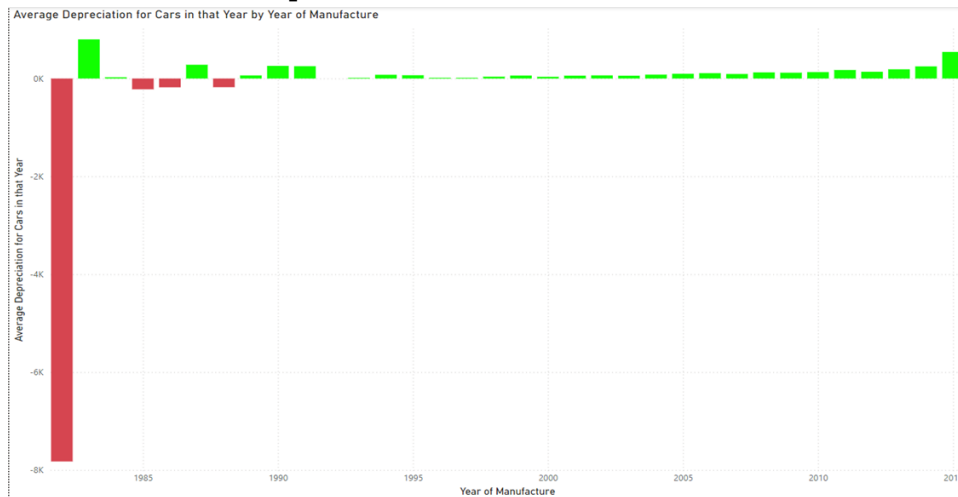
c) Insight 3

How does the year of manufacture influence the depreciation of car prices?

SQL query:

```
SELECT year, AVG(mmr - sellingprice) AS average_depreciation
FROM vehicle_sales
GROUP BY year
ORDER BY year;
```

Visualization of Output:

Average Depreciation for Cars in that Year by Year of Manufacture

Interpretation of Result:
Older cars, such as those manufactured before 1985, show significant depreciation in value, while cars from 2000 onwards maintain more stable prices, indicating higher retention of market value for newer models.

d) Insight 4

What is the impact of a car's condition on its price across different make and model categories?

SQL query:

```sql
SELECT make, model, `condition`, AVG(sellingprice) AS average_price
FROM vehicle_sales
WHERE `condition` IS NOT NULL and make IS NOT NULL AND TRIM(model) <> '' and model IS NOT
NULL AND TRIM(model) <> ''
GROUP BY make, model, `condition`
ORDER BY make, model, `condition` ;
```

Visualization of Output:
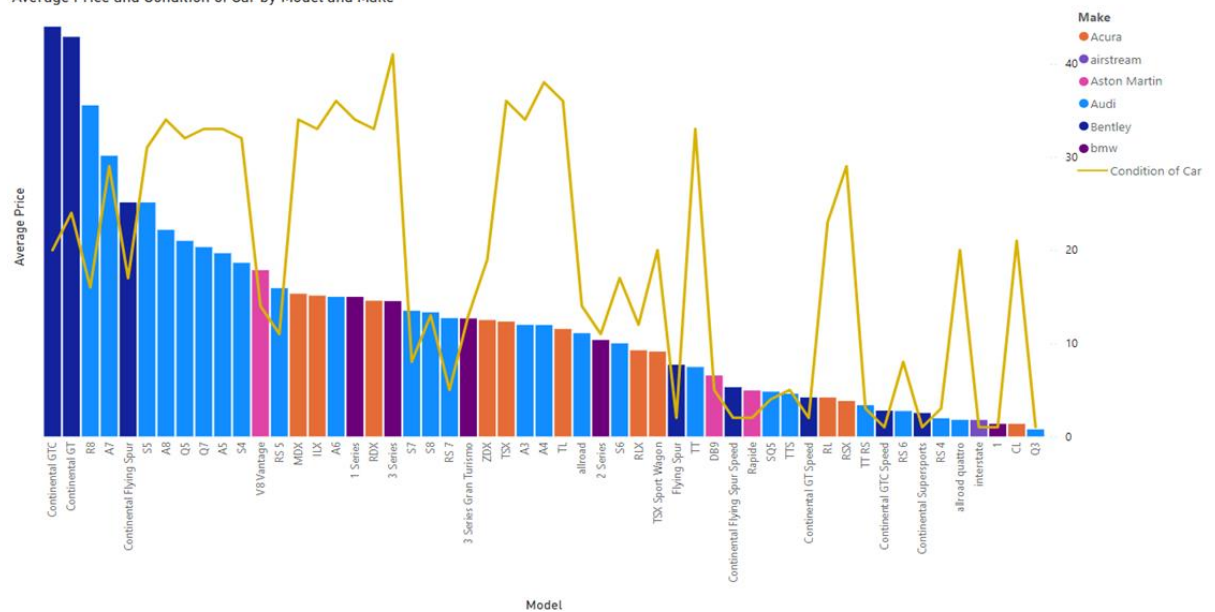
Interpretation of Result:
Models like the Bentley Continental GTC and Audi R8, despite varying conditions, command high average prices, indicating that luxury and brand reputation significantly outweigh condition in determining value.

e) Insight 5

What is the distribution of car prices in different regions, and are there notable regional price differences?

SQL query:

```
For Buyers- Identify regions with affordable pricing (low average price)
SELECT state, AVG(sellingprice) AS Avg_price

FROM vehicle_sales

GROUP BY state

ORDER BY AVG(sellingprice) asc;



For Sellers: Target regions with higher average prices for better profits.

SELECT state, AVG(sellingprice) AS Avg_price

FROM vehicle_sales

GROUP BY state

ORDER BY AVG(sellingprice) desc;
```
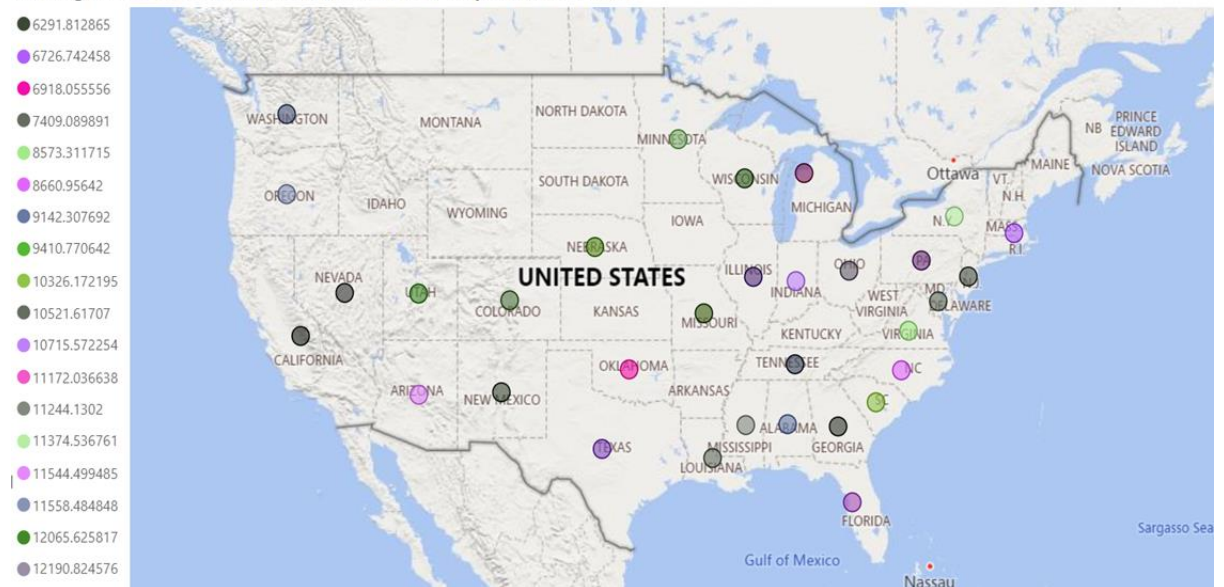
Visualization of Output:

Average Prices of Cars in Various States for Buyers and Sellers

Interpretation of Result:
States like California and Florida exhibit higher average car prices, making them attractive for sellers, while states like Michigan and Alabama show lower average prices, presenting opportunities for buyers seeking deals.
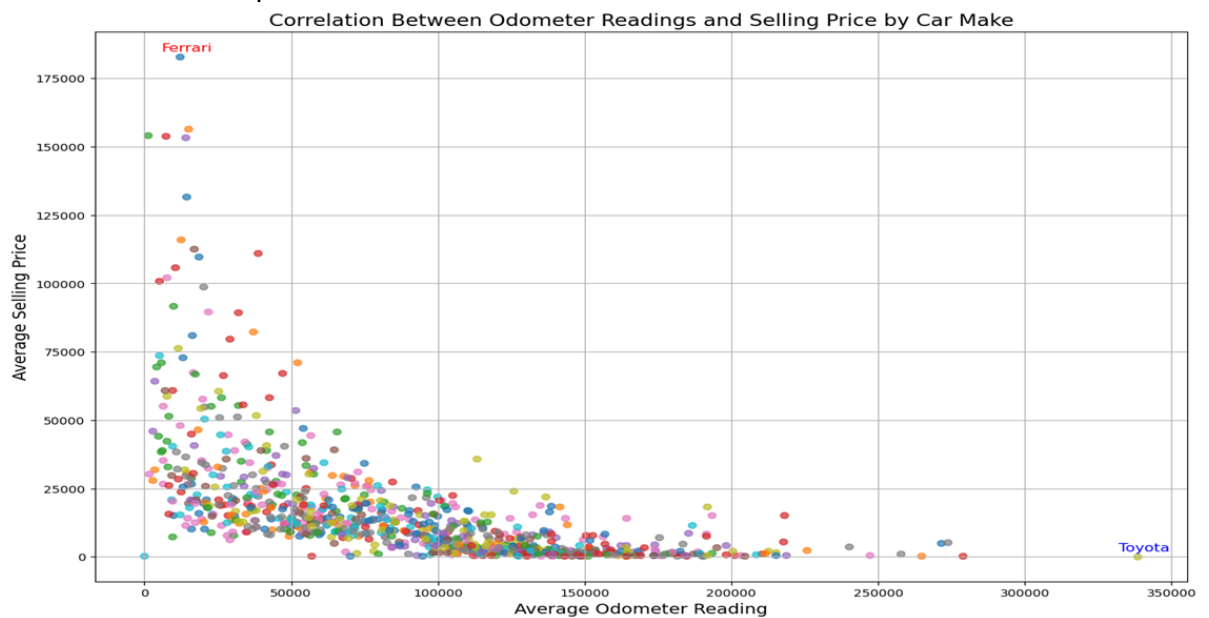
f) Insight 6

How do odometer readings correlate with selling price across various car makes and models?

SQL query:

```
SELECT make, model,
Avg(odometer) AS odometer_readings,
Avg(sellingprice) AS Avg_price
from vehicle_sales
Group by make, model
order by Avg_price DESC;
```

Visualization of Output:

Correlation Between Odometer Readings and Selling Price by Car Make

Interpretation of Result:
Luxury brands like Ferrari retain high prices regardless of mileage, while Toyota sees significant price drops with higher mileage, showing greater sensitivity to odometer readings.
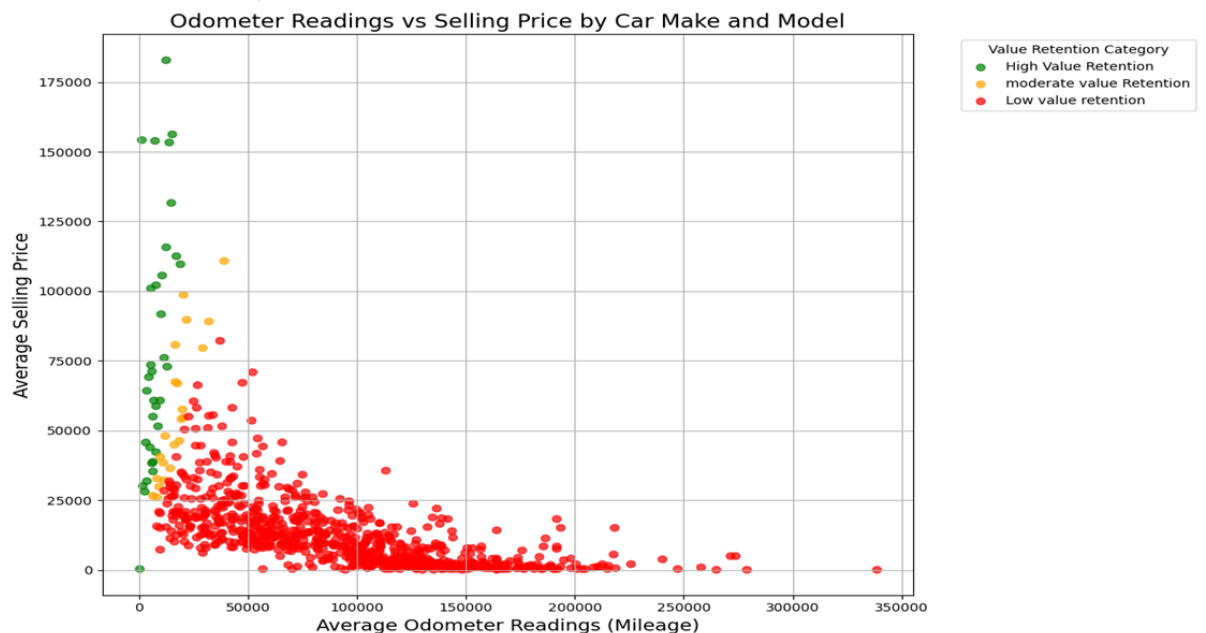
g) Insight 7

To correlate odometer readings with price and understand how mileage affects a car's value.

SQL query:

```
SELECT make, model,
Avg(odometer) AS Avg_odometer_readings,
Avg(sellingprice) AS Avg_price,
Count(*) AS Total_sales,
CASE
when Avg(sellingprice)/Avg(odometer) > 30 THEN 'High Value Retention'
when Avg(sellingprice) /Avg(odometer) Between 10 AND 30 Then 'moderate value Retention'
Else 'Low value retention'
END AS Value_Category
from vehicle_sales
group by make, model
order by Avg_price DESC;
```

Visualization of Output:

Interpretation of Result:
Luxury cars with high prices, such as Ferraris, fall into the high-value retention category, while most economy models show low retention, reflecting significant depreciation with increased mileage.
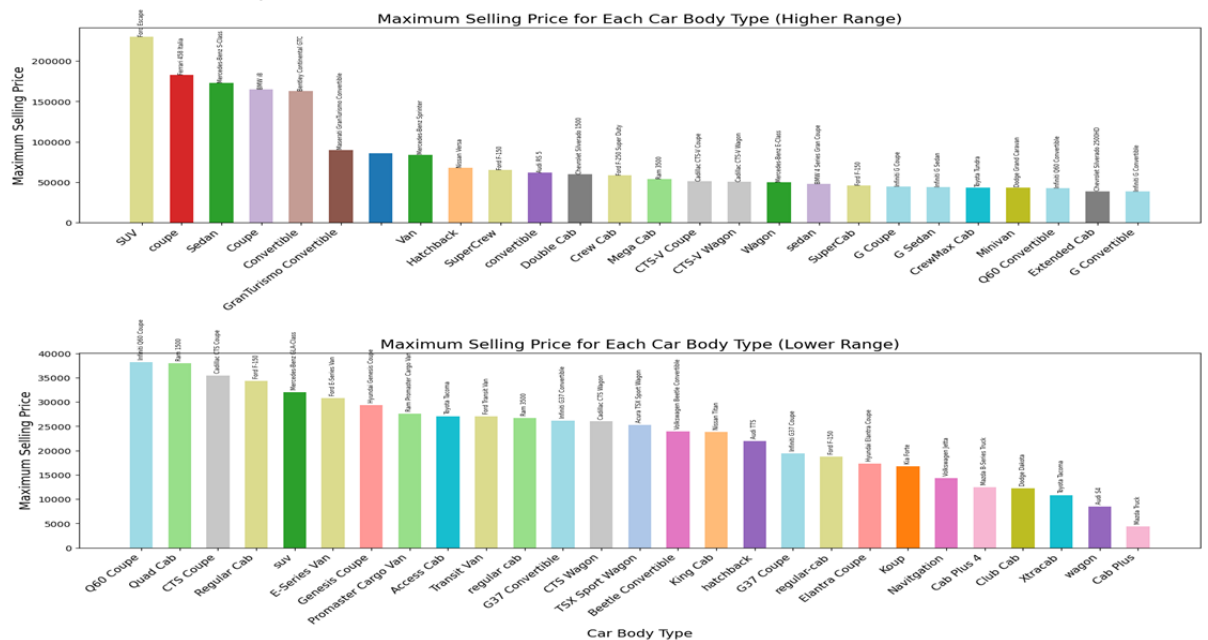
h) Insight 8

What are the most expensive cars sold by body type and their average market performance?

SQL query:

```
SELECT body, max(sellingprice) AS Max_price , make, model
from vehicle_sales
Group by body,make, model
Order by Max_price DESC;
```

Visualization of Output:

Interpretation of Result:

The top range, such as SUVs, Coupes, and Sedans, showcases the highest maximum selling prices, driven by luxury models. In contrast, the lower range includes more affordable categories like regular cabs and hatchbacks.
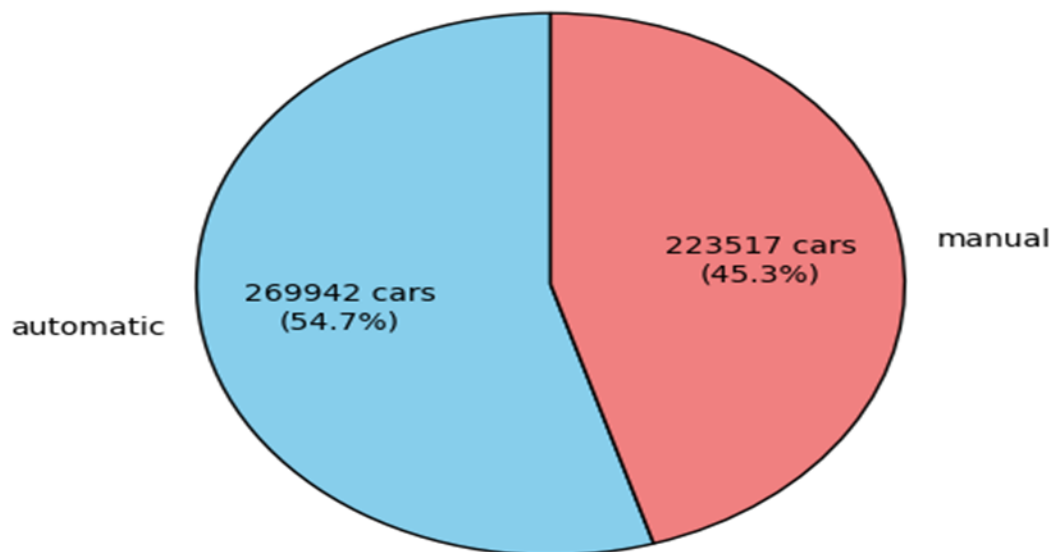
i) Insight 9

Which transmission types (automatic/manual) tend to fetch higher prices in the used car market?

SQL query:

```
SELECT transmission, Avg(sellingprice) as Avg_price, Count(*) as car_count
from vehicle_sales
where transmission IN ('automatic', 'manual')
Group by transmission
order by Avg_price DESC;
```

Visualization of Output:

Average Selling Price by Transmission Type

269942 cars
(54.7%)
automatic

223517 cars
(45.3%)
manual

Interpretation of Result:
Automatic transmission cars dominate the market with 54.7% of sales and tend to fetch slightly higher prices on average compared to manual transmission cars, which make up 45.3% of the market.
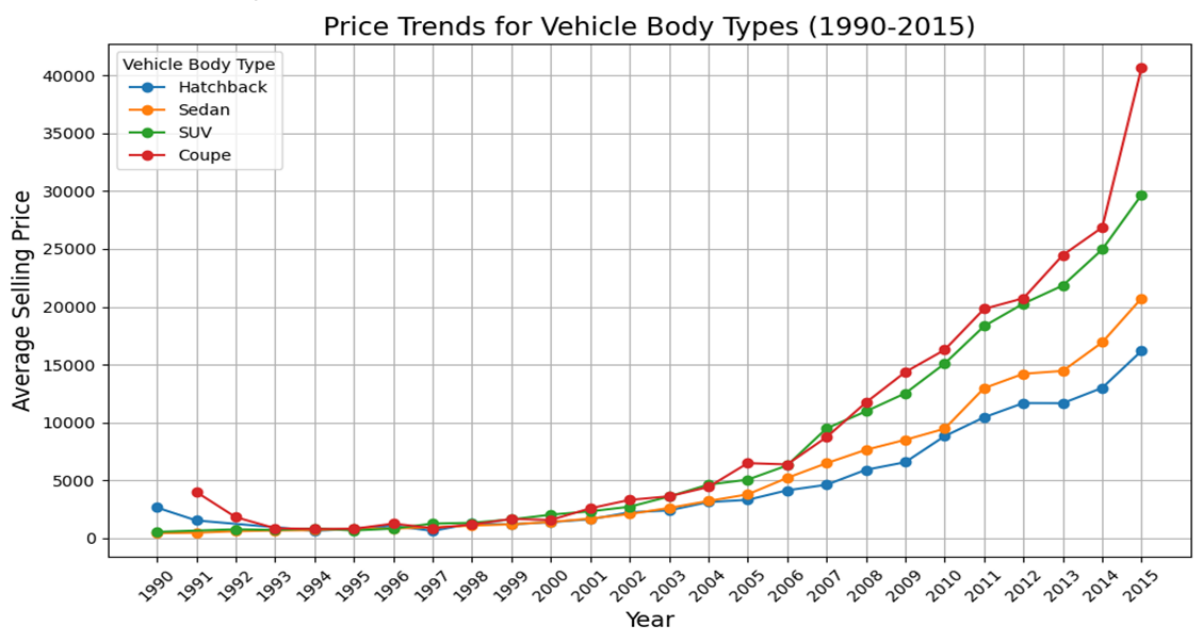
j) Insight 10

What are the price trends for specific vehicle body types (e.g., Hatchback, Sedan, Coupe, SUV) across different years?

SQL query:

```
SELECT year, body, avg(sellingprice)
from vehicle_sales
where body in ("Hatchback","SUV","Sedan","Coupe")
group by year, body
order by year,body;
```

Visualization of Output:

Price Trends for Vehicle Body Types (1990-2015)

Interpretation of Result:
From 1990 to 2015, Coupes exhibit the steepest price growth, reflecting their luxury appeal, while SUVs steadily increase in value. Sedans and Hatchbacks meanwhile follow a moderate price growth trajectory.
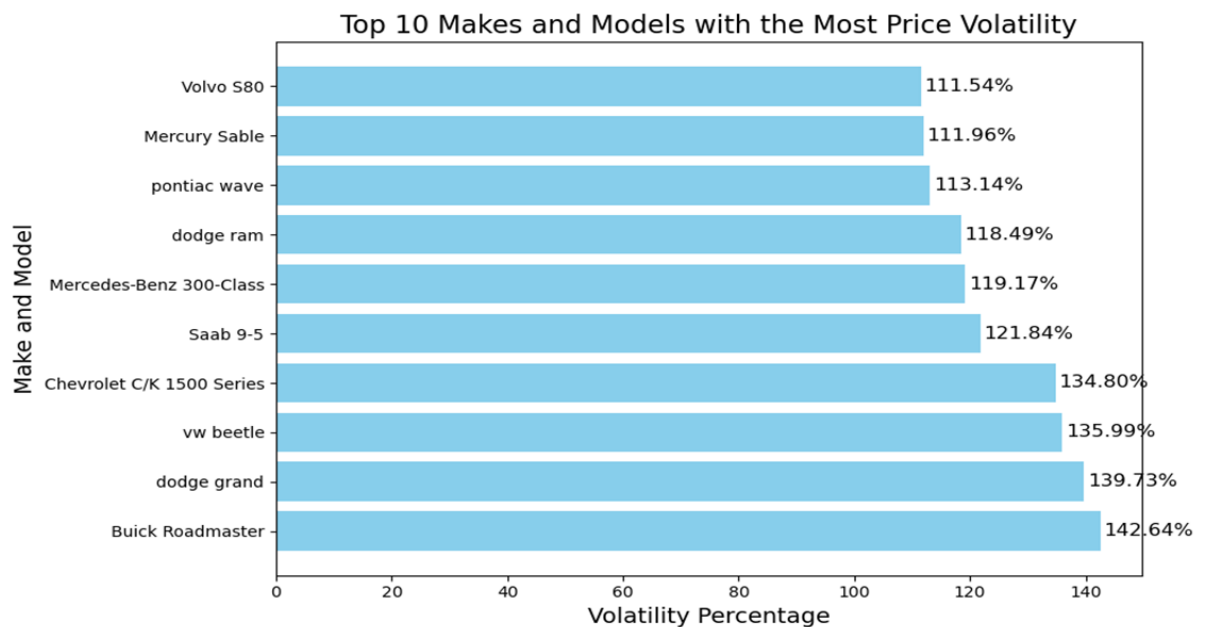
k) Insight 11

Identify Makes and Models with the Most Price Volatility

SQL query:

```
WITH price_stats AS (
  SELECT
    make,
    model,
    STDDEV_SAMP(sellingprice) AS price_std_dev,
    AVG(sellingprice) AS average_price
  FROM vehicle_sales
  WHERE
    sellingprice IS NOT NULL
    AND make IS NOT NULL AND TRIM(make) <> ''
    AND model IS NOT NULL AND TRIM(model) <> ''
  GROUP BY make, model
)
SELECT
  make,
  model,
  price_std_dev,
  average_price,
  (price_std_dev / average_price) * 100 AS volatility_percentage
FROM price_stats
```

Visualization of Output:

Top 10 Makes and Models with the Most Price Volatility

Interpretation of Result:
The Buick Roadmaster tops the list with the highest price volatility, followed by the Dodge Grand and VW Beetle, indicating significant fluctuations in their market prices, possibly due to demand.

l) Insight 12

Detecting Price Outliers by Make and Model

SQL query:

```
WITH ranked_prices AS (
  SELECT
    make,
    model,
    sellingprice,
    NTILE(100) OVER (PARTITION BY make, model ORDER BY sellingprice) AS percentile_rank
  FROM vehicle_sales
  WHERE
    make IS NOT NULL AND TRIM(make) <> ''
    AND model IS NOT NULL AND TRIM(model) <> ''
    AND sellingprice IS NOT NULL
)
SELECT
  make,
  model,
  sellingprice
FROM ranked_prices
WHERE percentile_rank > 95;
```

Visualization of Output:

Price Distribution by Model
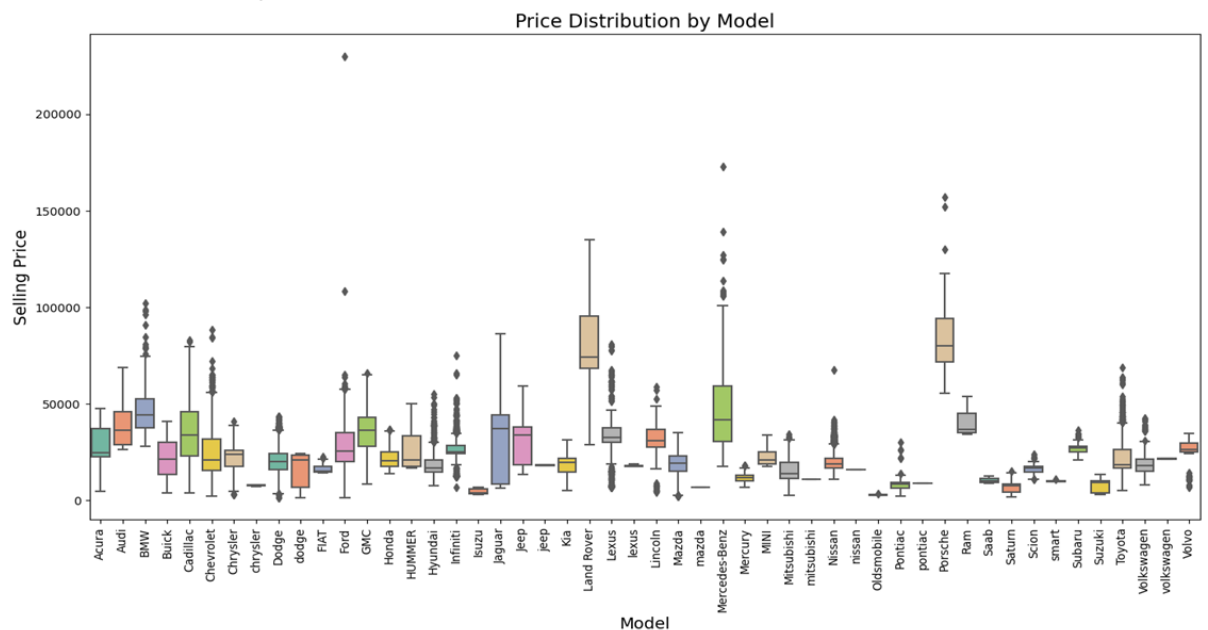
Interpretation of Result:
The plot reveals significant price outliers for luxury brands like Mercedes-Benz
and Porsche, reflecting rare high-value transactions. Economy brands show
narrower distributions, indicating more consistent pricing.

m)   Insight 13

How do car sales vary across quarters in different years?

SQL query:

```
SELECT
 EXTRACT(year FROM saledate) AS year,
 count(CASE WHEN EXTRACT(quarter FROM saledate) = 1
 THEN 1 END) AS Q1,
 count(CASE WHEN EXTRACT(quarter FROM saledate) = 2
 THEN 1 END) AS Q2,
 count(CASE WHEN EXTRACT(quarter FROM saledate) = 3
 THEN 1 END) AS Q3,
 count(CASE WHEN EXTRACT(quarter FROM saledate) = 4
 THEN 1 END) AS Q4
FROM vehicle_sales
where EXTRACT(year FROM saledate) is not null
GROUP BY EXTRACT(year FROM saledate)
ORDER BY EXTRACT(year FROM saledate);
```

Visualization of Output:

Car Sales Distribution Across Quarters for 2014 and 2015
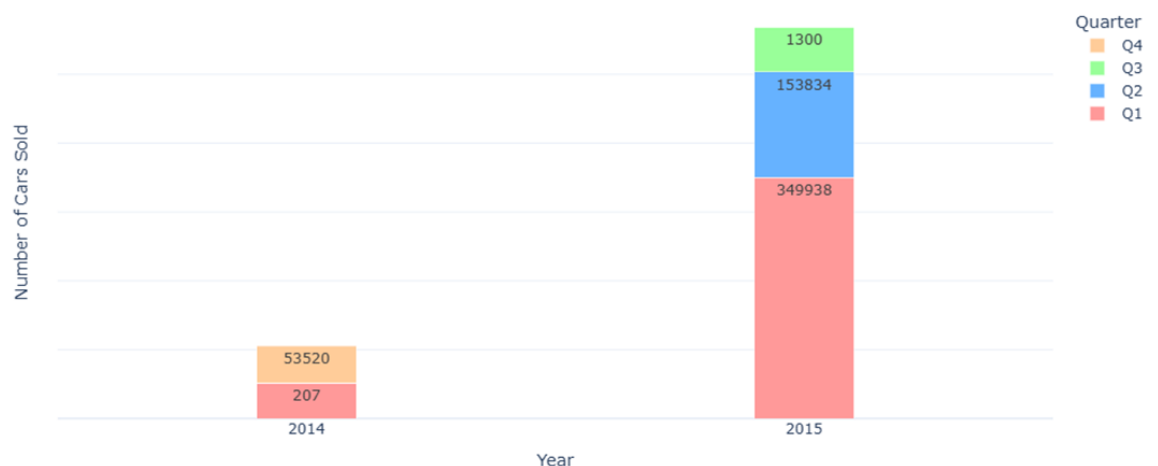
Interpretation of Result:
Car sales in 2015 saw a substantial increase compared to 2014, with Q1 leading
sales by a significant margin. Q2 and Q3 also showed strong performance, while
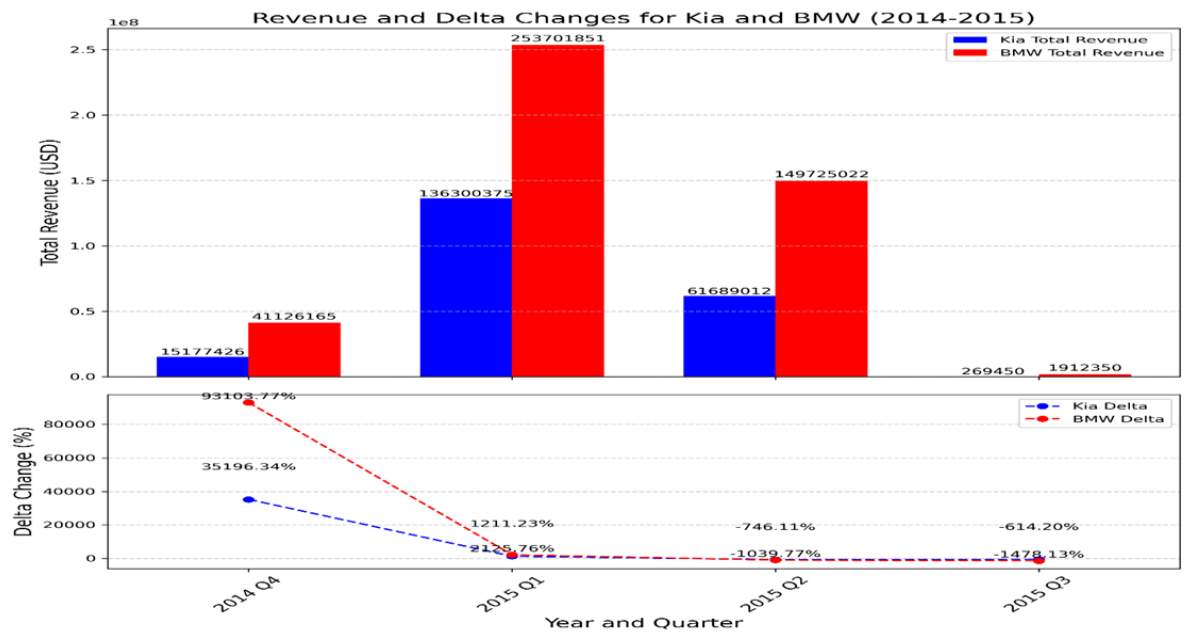Q4 remained relatively low across both years.

n) Insight 14

What is the percentage change in total sales revenue for Kia and BMW between 2014 and 2015 across quarters?

SQL query:

```
SELECT
 EXTRACT(year FROM saledate) AS revenue_year,
 EXTRACT(quarter FROM saledate) AS revenue_quarter,
 sum(case when make = "kia" then sellingprice else 0 end) AS total_revenue_kia,
 ROUND(
  100 * (
    sum(case when make = "kia" then sellingprice else 0 end) - LAG( sum(case when make = "kia"
then sellingprice else 0 end), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)))/
   cast(LAG(sum(case when make = "kia" then sellingprice else 0 end), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)) as decimal(10,3)),3
     )
 AS delta_kia,
 sum(case when make = "BMW" then sellingprice else 0 end) AS total_revenue_bmw,
 ROUND(
  100 * (
    sum(case when make = "BMW" then sellingprice else 0 end) - LAG( sum(case when make =
"BMW" then sellingprice else 0 end), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)))/
   cast(LAG(sum(case when make = "BMW" then sellingprice else 0 end), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)) as decimal(10,3)),3
     )
 AS delta_bmw
FROM vehicle_sales
where EXTRACT(year FROM saledate) is not null
GROUP BY
 EXTRACT(year FROM saledate),
 EXTRACT(quarter FROM saledate)
ORDER BY
 EXTRACT(year FROM saledate),
 EXTRACT(quarter FROM saledate);
```

Visualization of Output:

Revenue and Delta Changes for Kia and BMW (2014-2015)

Interpretation of Result:
Kia's revenue surged in Q4 2014 and Q1 2015 but declined sharply afterward.
BMW's revenue peaked in Q1 2015, followed by significant drops in Q2 and Q3,
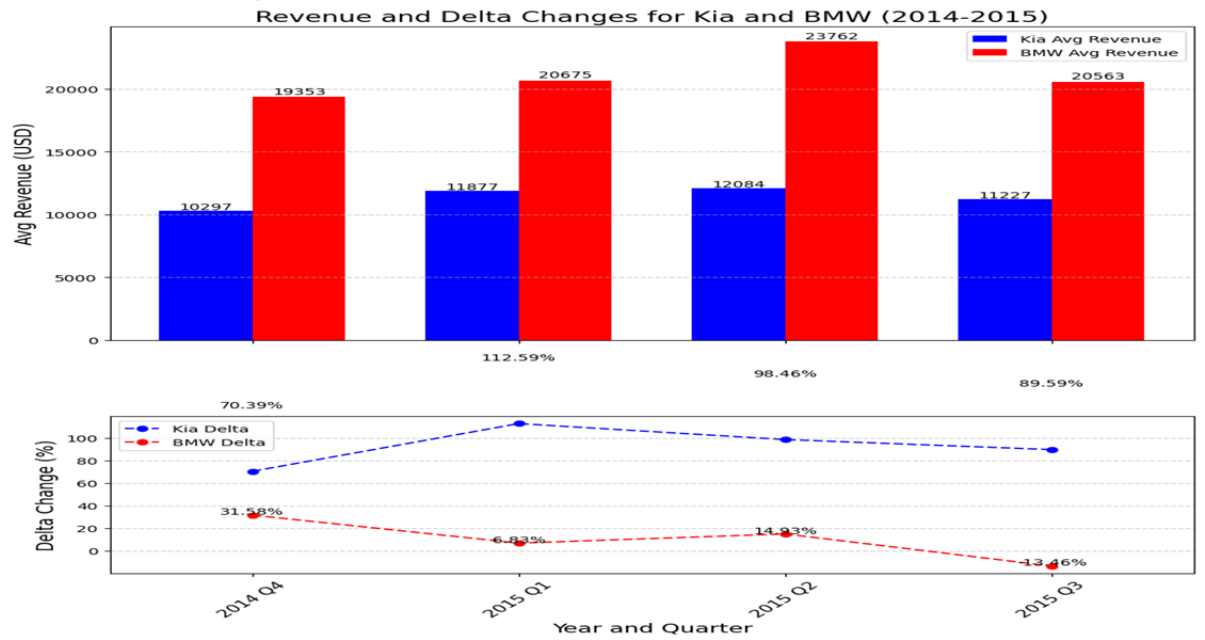reflecting market fluctuations.

o) Insight 15

What is the percentage change in average selling price for Kia and BMW between 2014 and 2015 across quarters?

SQL query:

```sql
SELECT
 EXTRACT(year FROM saledate) AS revenue_year,
 EXTRACT(quarter FROM saledate) AS revenue_quarter,
 cast( sum(case when make = "kia" then sellingprice else 0 end) / count(case when make = "kia" then
1 end) as
 decimal(10,3)) AS avg_revenue_kia,
 ROUND(
  100 * (
    cast( sum(case when make = "kia" then sellingprice else 0 end) / count(case when make = "kia"
then 1 end) as
 decimal(10,3)) - LAG( cast( sum(case when make = "kia" then sellingprice else 0 end) / count(case
when make = "kia" then 1 else 0 end) as
 decimal(10,3)), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)))/
  cast(LAG( cast( sum(case when make = "kia" then sellingprice else 0 end) / count(case when make
= "kia" then 1 end) as
 decimal(10,3)), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)) as decimal(10,3)),3
    )
 AS delta_kia,
 cast( sum(case when make = "bmw" then sellingprice else 0 end) / count(case when make = "bmw"
then 1 end) as
 decimal(10,3)) as avg_revenue_bmw,
 ROUND(
  100 * (
    cast( sum(case when make = "bmw" then sellingprice else 0 end) / count(case when make =
"bmw" then 1 end) as
 decimal(10,3)) - LAG( cast( sum(case when make = "bmw" then sellingprice else 0 end) /
count(case when make = "bmw" then 1 end) as
 decimal(10,3)), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)))/
  cast(LAG( cast( sum(case when make = "bmw" then sellingprice else 0 end) / count(case when
make = "bmw" then 1 end) as
 decimal(10,3)), 1) OVER
 (ORDER BY EXTRACT(year FROM saledate), EXTRACT(quarter FROM saledate)) as decimal(10,3)),3
    )
 AS delta_bmw
FROM vehicle_sales
where EXTRACT(year FROM saledate) is not null
GROUP BY
 EXTRACT(year FROM saledate),
 EXTRACT(quarter FROM saledate)
ORDER BY
 EXTRACT(year FROM saledate),
 EXTRACT(quarter FROM saledate);
```

Visualization of Output:

Revenue and Delta Changes for Kia and BMW (2014-2015)

Interpretation of Result:
Kia's average selling price rose 70.39% in Q4 2014, grew further in Q1 2015,
then stabilized. BMW's price peaked in Q1 2015 before declining in later quarters.

V. Conclusion

By exploring various aspects of the data, we have identified key trends and actionable insights that can inform decision-making for stakeholders in the automotive industry. Below are the findings and their implications:

•Certain car brands have shown significant price appreciation over time, indicating strong market demand and brand value retention.

•Regional analysis highlighted significant price differences, revealing affordable markets for buyers and profitable markets for sellers.

•Variables such as vehicle condition, transmission type, and odometer readings have a measurable impact on selling prices.

•Yearly and quarterly analyses showcased seasonal sales patterns and changing preferences, enabling better inventory planning and marketing schedules.

•Price and revenue trends for brands like Kia and BMW across specific years provided a comparative view of brand performance.

•For Dealers: Focus on high-value models, optimize inventory for peak sales periods, and target profitable regions.

•For Buyers: Identify affordable regions and high-value brands or models.

Thank You